形式言語の変換に関する研究

著者	小島 政明
号	364
発行年	1972
URL	http://hdl.handle.net/10097/9100

氏名（本籍）	小島政明（宮城県）
学位の種類	工学博士
学位記番号	工博第３６４号
学位授与年月日	昭和４７年１１月１日
学位授与の要件	学位規則第５条第１項該当
研究科専門課程	第北大学大学院工学研究科 （博士課程）電気及通信工学専攻
学位論文題目	形式言語の変換に関する研究
論文審査委員	（主査） 教授　本多　波雄　教授　大泉　充郎 教授　木村　正行　教授　野口　正一 助教授　那須　正和

論文内容要旨

形式言語の変換の研究は，１９５０年代後半から広く行われている形式言語の研究の一分野であり，プログラミング言語のコンパイリングなどの研究の基礎となるものである。形式言語の変換の問題をとり扱うときには，その言語を生成する形式文法の導出図の集合が特に重要な役割を果す。なぜなら，導出図は与えられた形式文法によって語がどのように生成されたかについての情報をすべて含んでいるからである。

本論文の主たる目的は，形成文法の導出図の集合の性質を解明することである。とり扱う形式文法は文脈自由形文法（ＣＦ文法と略記する）と文脈規定形文法（ＣＳ文法と略記する）である。ＣＦ文法の導出図がグラフ理論における木をなすことはよく知られている。さらに，本論文では，ＣＳ文法の導出図から導出木を定義する。よって，これらの導出図の集合をとり扱うためには，

木を入力とするオートマトンを定義することが必要である。

従来，木を入力とするオートマトン（いわゆる木オートマトン）は，多くの研究者によっていろいろな形で定義され，その性質が調べられている。特に，それらのオートマトンの性質は有限オートマトンの性質と類似していることが知られている。また，Thatcher が定義した木オートマトンはCF文法の導出図の集合を特性づける能力をもっている。

本論文においては，文脈規定形木オートマトン（STオートマトンと略記する）と呼ぶオートマトンおよび種々の木の変換装置（トランスジューサ）を定義し，従来ほとんど知られていなかったCS文法の導出図の集合を明らかにする。また，CF文法の導出図の集合を特性づける能力をもつ有限指定木オートマトンと呼ぶオートマトンを定義し，それが受理する木の集合の性質を調べることにより，いわゆる正規集合の概念を興味深い形で木の集合の上に拡張することができることを示す。さらに，階層的アルファベットの上で木トランジションシステムを定義し，その性質を調べることにより，前述した有限オートマトンと木オートマトンとの類似性を一層明らかにする。

得られる諸結果は，木の集合のとり扱いを容易にし，CS文法およびCF文法によって生成される形式言語の構文解析やトランスレーションの研究に有用な手段を与えるものである。

以下に，本論文において議論する内容を章別に簡単に説明し，得られる結果の主なものをあげる。

第１章　基本的定義と性質

第２章以下において種々のオートマトンおよびトランスジューサを定義するが，それらの入力となるべき木の定義をまず述べる。次に，形式言語について述べる。特に，CS文法の導出図および導出木について詳述する。さらに，CS文法の導出の標準形について述べる。ここでは，記法について簡単に説明するにとどめる。

Σを有限アルファベットとするとき，Σの上すべての木の集合をT_Σで表わし，T_Σのすべての部分集合のクラスを$\mathcal{T}_\Sigma$で表わす。文法（CS文法およびCF文法）をGで表わし,Gが生成する言語をL(G)で表わす。また，文法Gの導出木の集合をD(G)で表わす。

第２章　文脈規定形木オートマトン

従来定義されている木オートマトンを拡張して，文脈規定形木オートマトン（STオートマトン）と呼ぶオートマトンを定義し，その性質を調べる。

（定義２．１．１）（註1）T_Σの上のSTオートマトンは４字組$K=<\Sigma, P, f, P_F>$である。ここに，Σは有限アルファベット，Pは状態の有限集合，fはΣから$P^* \times P^* \times P^* \times P$有限部分集合のクラスへの写像，$P_F$はPの部分集合で最終状態の集合である。ただし，$f(\sigma) \ni$

$(\chi_1, \chi_2, \chi_3, p)$ かつ $\chi_2 = \lambda$（註2）ならば $\chi_1 = \chi_3 = \lambda$ である。

ＳＴオートマトンＫへの入力はT_Σの要素，つまりΣの上の木である。ＳＴオートマトンＫによって受理される木の集合をA(K)とかく。（ＳＴオートマトンの動作についてはここでは省略する。）木の集合ＴがあるＳＴオートマトンＫについてＴ＝A(K)となるとき，ＴをＳＴオートマトンによって受理可能であるという。

ＳＴオートマトンによって受理可能なT_Σの部分集合のクラスを$\mathcal{T}_{\Sigma,ST}$で表わす。

まず，$\mathcal{T}_{\Sigma,ST}$の閉包性質についての結果をあげる。

〔定理２．２．１〕　和集合の演算で閉じている

〔定理２．２．２〕　補集合の演算では閉じていない。

〔定理２．２．３〕　$\mathcal{T}_{\Sigma,TA}$（註3）に属する集合との共通集合の演算で閉じている。

〔定理２．２．５〕　木トランスジューサ写像で閉じている。

〔系　２．２．６〕　プロジェクションで閉じている。ここにプロジェクションとは木の名前のつけかえを指定する写像である。

次に，ＳＴオートマトンはＣＳ文法の導出図の集合を特性づける能力をもっていることを示す。

通常のＣＳ文法を少し変形してＭＣＳ文法と呼ぶ文法を定義する。（註4）ＭＣＳ文法の生成能力は通常のＣＳ文法の生成能力と同じである。

〔補題２．３．２〕　ＧをＭＣＳ文法とする。このとき，あるＳＴオートマトンＫが存在してD(G)＝A(K)である。

〔補題２．３．３〕　ＫをＳＴオートマトンとする。このとき，あるＭＣＳ文法Ｇとプロジェクションτが存在してA(K)＝ τ（D(G)）である。

これらの補題および系２．２．６から次の定理を得る。

〔定理２．３．４〕　木の集合は，それがあるＭＣＳ文法の導出木のプロジェクションであるとき，かつそのときに限り，ＳＴオートマトンによって受理可能である。

次に，ＳＴオートマトンの受理能力についての結果をあげる。

〔定理２．４．７〕　$\mathcal{T}_{\Sigma,ST}$はT_Σの帰納的部分集合のクラスの真の部分クラスである。

〔定理２．４．１２〕$\mathcal{T}_{\Sigma,TA}$は$\mathcal{T}_{\Sigma,ST}$の真の部分クラスである。

この章の最後に，ＳＴオートマトンのハイアラーキについて述べてあるが，ここでは省略する。

（註１）　定義，定理などの番号は本論文の番号と同じである。例えば，定義２．１．１．は第２章・第１節の１番目の定義である。

（註２）　λは空系列を表わす。

（註３）　$\mathcal{T}_{\Sigma,TA}$については第３章参照。

（註４）　ＭＣＳ文法Ｇについても，その導出木の集合をD(G)とかく。

なお，STオートマトンはCS言語の受理装置として知られている線形拘束オートマトンなどと比較して，そのとり扱いが非常に容易であるので，この章で与えた結果は，従来あまり研究が進んでいなかったCS言語の理論に貢献するものと思う。

第3章　有限指定木オートマトン

2重名前付木およびその集合の上の三つの演算（和，木連接および木スター）を定義し，それらの演算を用いて，有限指定木オートマトンと呼ぶオートマトンによって受理される木の集合のクラスの特性化を行う。

（定義3.1.1）　T_Δ の上の有限指定木オートマトン（木オートマトンと略記する）は4字組 $M=\langle \Delta, P, f, P_F \rangle$ である。ここに，Δ は有限アルファベット，Pは状態の有限集合，fは Δ から $P^* \times P$ の有限部分集合のクラスへの写像，P_F はPの部分集合で最終状態の集合である。

木オートマタはSTオートマタの部分クラスである。また，木オートマトンによって受理可能な T_Δ の部分集合のクラスを $\mathcal{T}_{\Delta,TA}$ で表わす。

次に，Σ を有限アルファベット，Qを可付番無限アルファベットとして，$Q \times \Sigma$ の上の2重名前付木を定義する。$Q \times \Sigma$ の上のすべての2重名前付木の集合を $T_{Q \times \Sigma}$ で表わし，$T_{Q \times \Sigma}$ のすべての部分集合のクラスを $\mathcal{T}_{Q \times \Sigma}$ で表わす。さらに，$T_1, T_2 \in \mathcal{T}_{Q \times \Sigma}$ について，和 $T_1 \cup T_2$，木連接 $T_1 \circ T_2$，木スター $T_1^{\circledast}$ を定義する。

（定義3.2.3）$\mathcal{T}_{Q \times \Sigma}$ の部分クラスで，すべての有限集合を含み，和，木連接および木スターの演算で閉じている最小のクラスを $\mathcal{C}_{Q \times \Sigma}$ で表わす。

（定義3.3.1）　2重名前の第2要素を改めて名前づけるプロジェクションを τ とし，$\mathcal{C}_\Sigma = \tau(\mathcal{C}_{Q \times \Sigma})$ と定義する。

得られる結果は $\mathcal{C}_\Sigma = \mathcal{T}_{\Sigma,TA}$，つまり

〔定理3.3.9〕　T_Σ の部分集合は，名前の第2要素に Σ の要素をもつ2重名前付木の集合のクラスですべての有限集合を含み，和，木連接および木スターの演算で閉じている最小クラスの τ によるプロジェクションに属しているとき，かつそのときに限り，有限指定木オートマトンによって受理可能である。

この結果は，有限オートマトンによって受理される語の集合の，和，連接およびスターの演算を用いた特性化の興味深い拡張になっている。

第4章　階層的アルファベットの上で定義される木トランジションシステム

階層的アルファベットの上の木およびその木を入力とするトランジションシステムを定義し，その性質を調べる。前述したように，木オートマトンの性質は有限オートマトンの性質と類似し

ているが，この類似性は木を階層的アルファベットの上で定義したときに特に顕著に現われる。この章で与える結果はその興味深い例である。

（定義 4．1．4）　$\Sigma = \{\sigma_1, \cdots, \sigma_k\}$ とし，$<\Sigma, \eta>$(註)を階層的アルファベットとする。$<\Sigma, \eta>$の上で定義される木トランジションシステムは $k+1$ 字組 $H = <Q, f_1, \cdots f_k>$ である。ここに，Qは状態の集合，f_j，$1 \leqslant j \leqslant k$，は$Q^{\eta(\sigma_j)}$からQへの関数である。

最終的に得られる結果は次の定理である。

〔定理 4．4．4〕　階層的アルファベットの上で定義される既約木トランジションシステムの同形代表のクラスは束をなす。ここに，同形代表とは互いに同形システムのクラスのかってな代表である。

第５章　文脈規定形木トランスジューサ

文脈規定形木トランスジューサ（ＳＴトランスジューサと略記する）と呼ぶ木の変換装置を定義し，その性質を決定問題を中心にして調べる。ＳＴトランスジューサは，簡単にいえば，最終状態を指定しないＳＴオートマトンに出力をつけ加えたものである。なお，これはＣＳ言語の変換の一つのモデルである。詳細はここでは省略する。

以上，本論文で議論した内容を簡単に述べたが，$\mathcal{T}_{\Sigma, ST}$ なるクラスが共通集合の演算で閉じているか否かは未解決である。さらに，ＳＴオートマトンは一般には非決定性であるが，決定性ＳＴオートマトンの定義および研究は今後に期待される課題である。

（註）　ηは階層づけを与える関数$\eta : \Sigma \to \{0, 1, 2 \cdots\}$である。

審 査 結 果 の 要 旨

形式言語の変換の研究は，電子計算機ソフトウエヤの重要問題であるプログラム言語のコンパイリングなどの基礎をなすものである。この変換は語が形式文法により生成される過程に基づいて行われ，生成過程は導出図または導出木というグラフによって表現される。それゆえ，形式言語の変換の研究は，導出図の集合の性質を究明することにあるといえる。しかし，このような研究はようやくその緒についたという段階であって，これまで文脈自由形文法について若干の結果が得られているにすぎない。

著者は，文脈自由形文法およびより一般的文法である文脈規定形文法について，導出図からつくられる導出木を処理するものとして，文脈規定形木オートマトンをはじめとする種々の木変換システムを定義して，その性質を調べ，さらにこれらのシステムを用いて，導出図の集合の基本的な性質を解明した。本論文はその研究成果をまとめたもので全編5章よりなる。

第1章では，本論文を通じて用いられる概念として，木とその集合，文脈規定形文法の導出図と導出木および導出図に一意的に対応する導出の標準形などの定義を述べている。

第2章では，文脈規定形木オートマトンというシステムを定義し，このオートマトンによって受理される木の集合性質を調べ，さらにこのオートマトンを用いて，文脈規定形文法の導出木の集合の種々の性質を明らかにしている。特に木の集合が文脈規定形文法の導出木の集合のプロジェクションであることと，それが文脈規定形木オートマトンによって受理可能であるということとは同等な命題であるという注目すべき結果を導いている。

第3章では，有限指定木オートマトンを定義し，このオートマトンに関する一つの興味ある結果を与えている。すなわち，2重名前付木の集合の上の和，木連接および木スターという三つの演算を定義し，この三つの演算に関する閉包性を用いて，有限指定木オートマトンによって受理される木の集合を特性づけている。この結果は，通常の有限オートマトンに関する基本的な性質を木オートマトンに拡張したものとなっており，すぐれた成果である。

第4章では，木トランジションシステムという木の変換システムを定義し，その性質を調べて，このシステムのクラスが束をなすという結果を導いている。

第5章では，文脈規定形言語の変換装置のモデルとして文脈規定形木トランスジューサと呼ぶ木の変換システムを定義し，その性質を調べている。

以上要するに，本論文はプログラム言語のコンパイリングの基礎となる形式言語の変換という問題について，種々の木変換システムを定義してその性質を調べ，さらにそれらのシステムを用いて文法の導出図の集合の性質を解明したものであって，情報工学の発展に寄与するところが少なくない。

よって，本論文は工学博士の学位論文として合格と認める。